You will use SPSS for Windows 10.0 to analyze data for your group projects. SPSS is available on PC computers in most campus computer labs. This handout lists some frequently used SPSS commands. Please see me for assistance on other procedures. I encourage groups to meet with me to discuss methods of analysis for their projects.

### SPLIT A DATA FILE INTO GROUPS

For splitting a file into groups based on one variable. This is often used for analysis of change over time.

**Example: Split a multiple year data file into separate survey years (YEAR) for analysis**

- Click on `data` → `split` → `organize output by groups`. Put `YEAR` into the `groups based on` box. Click on `OK` to temporarily separate subsequent analysis by survey year.

**NOTE:** This procedure should precede trend analysis. For example, if you want to look at how the relationship between education and occupational prestige has changed between 1972 and 1996, perform this procedure before analyzing scatterplots and regression statistics.

### FREQUENCY AND PERCENTAGE DISTRIBUTIONS

For univariate frequency and percentage distributions, with graph(s) and measures of center and spread

**Example 1: Education (EDUC), a quantitative variable**

- Click on `analyze` → `descriptive statistics` → `frequencies` and place `EDUC` in the `variable(s)` box. In the frequencies window, click on the `statistics` button and add the appropriate measures of center and spread (mean, median, standard deviation, quartiles, etc.) to the list of statistics generated. Click on `continue`. In the frequencies window, click on the `charts` button and choose the desired graph (histogram). `Continue` → `OK` to construct the distribution and graphs and to calculate the statistics.

**Example 2: Attitudes toward the legalization of marijuana (GRASS), a categorical variable**

- Click on `analyze` → `descriptive statistics` → `frequencies` and place `GRASS` in the `variable(s)` box. In the frequencies window, click on the `statistics` button and add the appropriate measures of center and spread (mode) to the list of statistics generated. Click on `continue`. In the frequencies window, click on the `charts` button and choose the desired graph (bar chart or pie chart (click on percentages)). `Continue` → `OK` to construct the distribution and graphs and to calculate the statistics.
**TWO-WAY / CROSSTABULATION TABLES**
For descriptive and inferential analysis of relationships between categorical variables

**Example: Two-way table of association between attitudes toward the legalization of marijuana (GRASS) and gender (SEX)**

Click on **analyze → descriptive statistics → crosstabs**. In the **crosstabs** window, put **GRASS** (the Y/dependent/response variable) in the **rows** of the table and **SEX** (the Y/independent/explanatory variable) in the **columns** of the table. At the bottom of the window, click on the **cells** button. In the cells window, choose **column percentages → continue → OK**. Click on the **statistics** button. In the statistics window, choose **chi-square → continue → OK** to generate a two-way table with column percentages in the cells and the chi-square test for the significance of the association.

**HYPOTHESIS TEST FOR DIFFERENCE BETWEEN TWO MEANS**
For the test for the difference between two means and a 95% confidence interval of the difference between two means

**Example: Comparison of mean education levels (EDUC) of men and women (SEX)**

Click on **analyze → compare means → independent-samples t test**. Put **EDUC** (the quantitative variable) in the **test variable(s)** box and **SEX** (the dichotomous categorical variable) in the **grouping variable** box. Click on the **define groups** box and put **1** in the **group 1** box (for male) and **2** in the **group 2** box (for female). **Continue → OK** to generate the test statistics, p-values, and 95% confidence intervals.

**COMPARISON OF TWO OR MORE MEANS**
For means and standard deviations of two or more groups

**Example: Comparison of mean ages (AGE) of those who consider themselves very happy, pretty happy, and not too happy (HAPPY)**

Click on **analyze → compare means → means**. Put **AGE** (the quantitative variable) in the **dependent list** box and **HAPPY** (the categorical variable) in the **independent list** box. **Continue → OK** to generate means, standard deviations, and sample sizes for each group.

**NOTE**: This is not a hypothesis test. This procedure will generate descriptive statistics to accompany the ANOVA procedure.
HYPOTHESIS TEST FOR DIFFERENCE BETWEEN MORE THAN TWO MEANS – ANALYSIS OF VARIANCE (ANOVA)

For the test for the difference between more than two means using one-way analysis of variance (ANOVA)

Example: Comparison of mean ages (AGE) of those who consider themselves very happy, pretty happy, and not too happy (HAPPY)

Click on analyze → compare means → one-way ANOVA. Put AGE (the quantitative variable) in the dependent list box and HAPPY (the categorical variable) in the factor list box. Continue → OK to generate the ANOVA table.

NOTE: This procedure does not list measures of center and spread for each group. Follow the Comparison of two or more means instructions to generate these statistics.

SCATTERPLOTS

To check for linearity in bivariate linear regression analysis

Example: Scatterplot of correlation between education (EDUC) and occupational prestige (PRESTG80)

Click on graph → scatter → simple → define. Place PRESTG80 (the Y/dependent/response variable) in the Y axis box and EDUC (the X/independent/explanatory variable) in the X axis box. Click on titles and add a title for the graphs. Continue → OK to construct the scatterplot.

LEAST-SQUARES LINEAR REGRESSION ANALYSIS

For the descriptive and inferential statistics for bivariate linear regression analysis, with Pearson’s r, the prediction equation (a and b), and \( r^2 \)

Example: Correlation between education (EDUC) and occupational prestige (PRESTG80)

Click on analyze → regression → linear. Put PRESTG80 (the Y/dependent/response variable) in the dependent box and EDUC (the X/independent/explanatory variable) in the independent box. Click on OK to generate the values of r, \( r^2 \), a, and b.

NOTE: This procedure assumes a linear correlation. Follow the Scatterplots instructions to generate a scatterplot for analysis of linearity.